Shellsort With Three Increments

by Svante Janson and Donald E. Knuth

Abstract. A perturbation technique can be used to simplify and sharpen A. C. Yao's theorems about the behavior of shellsort with increments $(h, g, 1)$. In particular, when $h = \Theta(n^{7/15})$ and $g = \Theta(h^{1/5})$, the average running time is $O(n^{23/15})$. The proof involves interesting properties of the inversions in random permutations that have been h-sorted and g-sorted.

Shellsort, also known as the "diminishing increment sort" [7, Algorithm 5.2.1D], puts the elements of an array $(X_0, \ldots, X_{n-1})$ into order by successively performing a straight insertion sort on larger and larger subarrays of equally spaced elements. The algorithm consists of t passes defined by increments $(h_{t-1}, \ldots, h_1, h_0)$, where $h_0 = 1$; the jth pass makes $X_k \le X_l$ whenever $l - k = h_{t-j}$.

A. C. Yao [11] has analyzed the average behavior of shellsort in the general three-pass case when the increments are $(h, g, 1)$. The most interesting part of his analysis dealt with the third pass, where the running time is $O(n)$ plus a term proportional to the average number of inversions that remain after a random permutation has been h-sorted and g-sorted. Yao proved that if g and h are relatively prime, the average number of inversions remaining is

$$\psi(h, g)n + \widehat{O}(n^{2/3}), \tag{0.1}$$

where the constant implied by $\widehat{O}$ depends on g and h. He gave a complicated triple sum for $\psi(h, g)$, which is too difficult to explain here; we will show that

$$\psi(h, g) = \frac{1}{2} \sum_{d=1}^{g-1} \sum_{r} \binom{h-1}{r} \left(\frac{d}{g}\right)^r \left(1 - \frac{d}{g}\right)^{h-1-r} \left| r - \left\lfloor \frac{hd}{g} \right\rfloor \right| . \tag{0.2}$$

Moreover, we will prove that the average number of inversions after such h-sorting and g-sorting is

$$\psi(h, g)n + O(g^3 h^2), \tag{0.3}$$

where the constant implied by O is independent of g, h, and n.

The main technique used in proving (0.3) is to consider a stochastic algorithm $\mathcal{A}$ whose output has the same distribution as the inversions of the third pass of shellsort. Then by slightly perturbing the probabilities that define $\mathcal{A}$, we will obtain an algorithm $\mathcal{A}^*$ whose output has the expected value $\psi(h, g)n$ exactly. Finally we will prove that the perturbations cause the expected value to change by at most $O(g^3 h^2)$.

Section 1 introduces basic techniques for inversion counting, and section 2 adapts those techniques to a random input model. Section 3 proves that the crucial random variables needed for inversion counting are nearly uniform; then section 4 shows that the leading term $\psi(h, g)n$ in (0.3) would be exact if those variables were perfectly uniform. Section 5 explains how to perturb them so that they are indeed uniform, and section 6 shows how this perturbation yields the error term $O(g^3h^2)$ of (0.3).

The asymptotic value of $\psi(h,g)$ is shown to be $(\pi h/128)^{1/2}g$ in section 7. The cost of the third pass in $(ch, cg, 1)$-shellsort for $c > 1$ is analyzed in section 8. This makes it possible to bound the total running time for all three passes, as shown in section 9, leading to an $O(n^{23/15})$ average running time when h and g are suitably chosen.

The bound $O(g^3h^2)$ in (0.3) may not be best possible. Section 10 discusses a conjectured improvement, consistent with computational experiments, which would reduce the average cost to $O(n^{3/2})$, if it could be proved.

The tantalizing prospect of extending the techniques of this paper to more than three increments is explored briefly in section 11.

1. Counting inversions. We shall assume throughout this paper that g and h are relatively prime. To fix the ideas, suppose $h = 5$, $g = 3$, $n = 20$, and suppose we are sorting the 2-digit numbers

$$(X_0, X_1, \ldots, X_{n-1}) = (03, 14, 15, 92, 65, 35, 89, 79, 32, 38, 46, 26, 43, 37, 31, 78, 50, 28, 84, 19)\,.$$

(Cf. [6, Eq. 3.3–(1)].) The first pass of shellsort, h-sorting, replaces this array by

$$(X'_0, X'_1, \ldots, X'_{n-1}) = (03, 14, 15, 32, 19, 35, 26, 28, 37, 31, 46, 50, 43, 84, 38, 78, 89, 79, 92, 65)\,.$$

The second pass, g-sorting, replaces it by

$$(X''_0, X''_1, \ldots, X''_{n-1}) = (03, 14, 15, 26, 19, 35, 31, 28, 37, 32, 46, 38, 43, 65, 50, 78, 84, 79, 92, 89)\,.$$

Our task is to study the inversions of this list, namely the pairs k, l for which $k < l$ and $X''_k > X''_l$.

The result of g-sorting is the creation of g ordered bits $X''_j < X''_{j+g} < X''_{j+2g} < \cdots$ for $0 \le j < g$, each of which contains no inversions within itself. So the inversions remaining are inversions between different sublists. For example, the 20 numbers sorted above lead to

$$\begin{aligned}
\text{list } 0 &= (03, 26, 31, 32, 43, 78, 92)\,,\\
\text{list } 1 &= (14, 19, 28, 46, 65, 84, 89)\,,\\
\text{list } 2 &= (15, 35, 37, 38, 50, 79)\,;
\end{aligned}$$

the inversions between list 0 and list 1 are the inversions of

$$(03, 14, 26, 19, 31, 28, 32, 46, 43, 65, 78, 84, 92, 89)\,.$$

It is well known [7, §5.21] that two interleaved ordered lists of lengths m have $\sum_{r=0}^{m-1} |r - s_r|$ inversions, where s_r of the elements of the second list are less than the $(r+1)$st element of the first list; for example, $(03, 14, 26, \ldots, 89)$ has

$$|0-0| + |1-2| + |2-3| + |3-3| + |4-3| + |5-5| + |6-7| = 4$$

inversions. If $r \geq s_r$, the $(r+1)$st element of the first list is inverted by $r - s_r$ elements of the second; otherwise it inverts $s_r - r$ of those elements. (We assume that the list elements are distinct.) The same formula holds for interleaved ordered lists of lengths m and $m-1$, because we can imagine an infinite element at the end of the second list.

Let Y_{kl} be the number of elements $X_{k'}$ such that $k' \equiv k \pmod h$ and $X_{k'} < X_l$. The n numbers Y_{ll} for $0 \leq l < n$ clearly characterize the permutation performed by h-sorting; and it is not hard to see that the full set of hn numbers Y_{kl} for $0 \leq k < h$ and $0 \leq l < n$ is enough to determine the relative order of all the X's.

There is a convenient way to enumerate the inversions that remain after g-sorting, using the numbers Y_{kl}. Indeed, let

$$J_{kl} = (k \bmod h + hY_{kl}) \bmod g\,. \tag{1.1}$$

Then X_l will appear in list $j = J_{ll}$ after g-sorting. Let S_{jl} be the number of elements $X_{k'}$ such that $X_{k'} < X_l$ and $X_{k'}$ is in list j. The inversions between lists j and j' depend on the difference $|S_{jl} - S_{j'l}|$ when X_l goes into list j.

Given any values of j and j' with $0 \leq j < j' < g$, let $j_s = (j + hs) \bmod g$, and let d be minimum with $j_d = j'$. Thus, d is the distance from j to j' if we count by steps of h modulo g. Let

$$H = \{j_1, j_2, \ldots, j_d\} \tag{1.2}$$

be the h numbers between j and j' in this counting process, and let Q_l be the number of indices k such that $0 \leq k < h$ and $J_{kl} \in H$. Then we can prove the following basic fact:

Lemma 1. *Using the notation above, we have*

$$S_{jl} - S_{j'l} = Q_l - \lfloor hd/g \rfloor \tag{1.3}$$

for all j, j', and l with $0 \leq j < j' < g$ and $0 \leq l < n$.

Proof. Since the X's are distinct, there is a permutation $(l_0, l_1, \ldots, l_{n-1})$ of $\{0, 1, \ldots, n-1\}$ such that $X_{l_0} < X_{l_1} < \cdots < X_{l_{n-1}}$. We will prove (1.3) for $l = l_t$ by induction on t.

Suppose first that $l = l_0$, so that X_l is the smallest element being sorted. Then $Y_{kl} = 0$ for all k, hence $J_{kl} = k \bmod g$ for $0 \leq k < h$. Also $S_{jl} = S_{j'l} = 0$. Therefore (1.3) is equivalent in this case to the assertion that *precisely $\lfloor hd/g \rfloor$ elements of the multiset*

$$\{0 \bmod g,\ 1 \bmod g,\ \ldots,\ (h-1) \bmod g\}$$

belong to H.

A clever proof of that assertion surely exists, but what is it? We can at any rate use brute force by assuming first that $j = 0$. Then the number of solutions to $x \equiv hd \pmod g$ and $0 \leq x < h$ is the number of integers in the interval $\bigl[-hd/g \,..\, -h(d-1)/g\bigr)$, namely $\lceil -h(d-1)/g \rceil - \lceil -hd/g \rceil = \lfloor hd/g \rfloor - \lfloor h(d-1)/g \rfloor$. Therefore the assertion for $j = 0$ follows by induction on d. And once we've proved it for some pair $j < j'$, we can prove it for $j+1 < j'+1$, assuming that $j'+1 < g$: The value of d stays the same, and the values of $j_1, j_2, \ldots, j_d$ increase by 1 (mod g). So we lose one solution if $j_s \equiv h-1 \pmod{g}$ for some s with $1 \le s \le d$; we gain one solution if $j_s \equiv -1 \pmod{g}$ for some s. Since $j_s \equiv h-1 \iff j_{s-1} \equiv -1$, the net change is zero unless $j_1 \equiv h-1$ (but then $j = g-1$) or $j_d \equiv -1$ (but then $j' = g-1$). This completes the proof by brute force when $l = l_0$.

Suppose (1.3) holds for $l = l_t$; we want to show that it also holds for when l is replaced by $l' = l_{t+1}$. The numbers Y_{kl} and $Y_{kl'}$ are identical for all but one value of k, since

$$Y_{kl'} = Y_{kl} + [l \equiv k \pmod{h}] \,.$$

Thus, the values of J_{kl} and $J_{kl'}$ are the same except that J_{kl} increases by h (mod g) when $k \equiv l \pmod{h}$. It follows that

$$Q_{l'} = Q_l + [J_{ll} = j] - [J_{ll} = j'] \,.$$

This completes the proof by induction on t, since $S_{jl'} = S_{jl} + [J_{ll} = j]$ for all j. $\square$

Corollary. *Using the notations above, the total number of inversions between lists j and j' is*

$$\sum_{l=0}^{n-1} \bigl|\, Q_l - \lfloor hd/g \rfloor \,\bigr|\, [J_{ll} = j] \,. \tag{1.4}$$

Proof. This is $|S_{jl} - S_{j'l}| = |r - s_r|$ summed over all r such that X_l is the $(r+1)$st element of list j. $\square$

In the example of $n = 20$ two-digit numbers given earlier, with $h = 5$, $g = 3$, $j = 0$, and $j' = 1$, we have $d = 2$, $H = \{2, 1\}$,

$l =$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$X_l =$	03	14	15	92	65	35	89	79	32	38	46	26	43	37	31	78	50	28	84	19
$Y_{0l} =$	0	1	1	4	3	1	4	4	1	2	2	1	2	2	1	3	3	1	4	1
$Y_{1l} =$	0	0	1	4	3	2	3	3	2	2	2	1	2	2	2	3	2	2	3	1
$Y_{2l} =$	0	0	0	4	3	2	4	3	2	2	3	1	2	2	2	3	3	1	4	1
$Y_{3l} =$	0	0	0	3	2	1	3	2	0	2	2	0	2	1	0	2	2	0	2	0
$Y_{4l} =$	0	0	0	4	3	2	4	4	2	2	3	1	3	2	1	4	3	1	4	0
$J_{0l} =$	$\underline{0}$	2	2	2	0	$\underline{2}$	2	2	2	1	$\underline{1}$	2	1	1	2	$\underline{0}$	0	2	2	2
$J_{1l} =$	1	$\underline{1}$	0	0	1	2	$\underline{1}$	1	2	2	2	$\underline{0}$	2	2	2	1	$\underline{2}$	2	1	0
$J_{2l} =$	2	2	$\underline{2}$	1	2	0	1	$\underline{2}$	0	0	2	1	$\underline{0}$	0	0	2	2	$\underline{1}$	1	1
$J_{3l} =$	0	0	0	$\underline{0}$	1	2	0	1	$\underline{0}$	1	1	0	1	$\underline{2}$	0	1	1	0	$\underline{1}$	0
$J_{4l} =$	1	1	1	0	$\underline{1}$	2	0	0	2	$\underline{2}$	1	0	1	2	$\underline{0}$	0	1	0	0	$\underline{1}$
$Q_l =$	3	4	3	2	4	4	3	4	3	4	5	2	4	4	2	3	4	3	4	3

and the underlined values J_{ll} are 0 for $l = 0, 3, 8, 11, 12, 14, 15$ (accounting for the seven elements in list 0). The inversions between lists 0 and 1 are therefore

$$|3-3| + |2-3| + |3-3| + |2-3| + |4-3| + |2-3| + |3-3| = 4$$

according to (1.4).

2. Random structures. We obtain a random run of shellsort if we assume that the input array $(X_0, X_1, \ldots, X_{n-1})$ is a random point in the n-dimensional unit cube. For each integer l in the range $0 \le l < n$ and for each "time" t in the range $0 \le t \le 1$ we will consider the contribution made by X_l to the total number of inversions if $X_l = t$.

Thus, instead of the quantities Y_{kl} and J_{kl} defined in the previous section, we define

$$Y_{kl}(t) = \sum_{\substack{k' \equiv k \pmod{h} \\ 0 \le k' < n}} [X_{k'} < t]\,, \tag{2.1}$$

$$J_{kl}(t) = \bigl(k \bmod h + hY_{kl}(t)\bigr) \mod g\,. \tag{2.2}$$

These equations are almost, but not quite, independent of l, because we assume that $X_l = t$ while all other X's are uniformly and independently random.

For each pair of indices j and j' with $0 \le j < j' < g$, we define H as in (1.2), and we let

$$Q_l(t) = \sum_{k=0}^{h-1} [J_{kl}(t) \in H][k \ne l \bmod h]\,. \tag{2.3}$$

This definition is slightly different from our original definition of Q_l, because we have excluded the term for $k = l \bmod h$. However, formula (1.4) remains valid because $j \notin H$; when $J_{ll} = j$, the excluded term is therefore zero.

Notice that, for fixed l, the random variables $Y_{kl}(t)$ for $0 \le k < h$ are independent. Therefore the random variables $J_{kl}(t)$ are independent; and $Q_l(t)$ is independent of $J_{ll}(t)$. The average contribution of X_l to the inversions between lists j and j' when $X_l = t$ is therefore

$$W_{jj'l}(t) = \Pr\,[J_{ll}(t) = j]\,\mathrm{E}\,\bigl|Q_l(t) - \lfloor hd/g \rfloor\bigr| \tag{2.4}$$

by (1.4), where probabilities and expectations are computed with respect to $(X_0, \ldots, X_{l-1}, X_{l+1}, \ldots, X_{n-1})$. The average total contribution of X_l is obtained by integrating over all values of t:

Lemma 2. *Let*

$$W_{jj'l} = \int_0^1 W_{jj'l}(t)\,dt\,. \tag{2.5}$$

Then the average grand total number of inversions in the third pass of shellsort is

$$\sum_{\substack{0 \le j < j' < g \\ 0 \le l < n}} W_{jj'l}\,. \quad \square \tag{2.6}$$

Our goal is to find the asymptotic value of this sum, by proving that it agrees with the estimate (0.3) stated in the introduction.

3. Near uniformity. The complicated formulas of the previous section become vastly simpler when we notice that each random variable $J_{kl}(t)$ is almost uniformly distributed: The probability that $J_{kl}(t) = j$ is very close to $1/g$, for each j, as long as t is not too close to 0 or 1. To prove this statement, it suffices to show that $Y_{kl}(t) \bmod g$ is approximately uniform, because h is relatively prime to g. Notice that $Y_{kl}(t)$ has a binomial distribution, because it is the sum of approximately n/h independent random 0–1 variables that take the value 1 with probability t.

Lemma 3. *If Y has the binomial distribution with parameters (m,t), then*

$$\left|\Pr\left[Y \bmod g = j\right] - \frac{1}{g}\right| < \frac{1}{g}\,\phi_{gm}(t) \tag{3.1}$$

for $0 \le j < g$, where

$$\phi_{gm}(t) = 2\sum_{k=1}^{\infty} e^{-8t(1-t)k^2 m/g^2}\,. \tag{3.2}$$

Proof. Let $y_j = \Pr\left[Y \bmod g = j\right]$, and consider the discrete Fourier transform

$$\hat{y}_k = \sum_{j=0}^{g-1} \omega^{kj} y_j = \mathrm{E}\,\omega^{kY}$$

where $\omega = e^{2\pi i/g}$. We have

$$\hat{y}_k = \sum_{l=0}^{m} \binom{m}{l} t^l (1-t)^{m-l}\omega^{kl} = (\omega^k t + 1 - t)^m\,, \tag{3.3}$$

and

$$\begin{aligned} |\omega^k t + 1 - t|^2 &= t^2 + (1-t)^2 + t(1-t)(\omega^k + \omega^{-k}) \\ &= 1 - 2t(1-t)(1 - \cos 2\pi k/g) \\ &= 1 - 4t(1-t)\sin^2 \pi k/g\,. \end{aligned} \tag{3.4}$$

If $0 \le x \le \pi/2$ we have $\sin x \ge 2x/\pi$; hence, if $0 \le k \le \frac{1}{2}g$,

$$|\omega^k t + 1 - t|^2 \le 1 - 16t(1-t)k^2/g^2 < e^{-16t(1-t)k^2/g^2}\,.$$

And if $\frac{1}{2}g < k < g$ we have $|\hat{y}_k| = |\hat{y}_{g-k}|$. Therefore

$$\sum_{k=1}^{g-1} |\hat{y}_k| \le 2\sum_{k=1}^{g/2} e^{-8t(1-t)k^2 m/g^2} < \phi_{gm}(t)\,. \tag{3.5}$$

The desired result follows since

$$y_j = \frac{1}{g}\sum_{k=0}^{g-1} \omega^{-kj}\hat{y}_k$$

and thus

$$\left|y_j - \frac{1}{g}\right| = \left|\frac{1}{g}\sum_{k=1}^{g-1}\omega^{-kj}\hat{y}_k\right| \le \frac{1}{g}\sum_{k=1}^{g-1}|\hat{y}_k|\,. \qquad \square$$

Corollary. *We have*

$$\left|\Pr\left[J_{kl}(t) = j\right] - \frac{1}{g}\right| < \frac{1}{g}\,\phi(t) \tag{3.6}$$

for $0 \le k < h$, where

$$\phi(t) = \begin{cases} 2\sum_{k=1}^{\infty} e^{-4t(1-t)k^2 n/g^2 h}\,, & \text{if } n \ge 4h; \\ g, & \text{if } n < 4h. \end{cases} \tag{3.7}$$

Proof. Each variable $Y_{kl}(t)$ in (2.1) for $0 \le k < h$ has the binomial distribution with parameters (m,t), where if $n \ge 4h$

$$m = \lceil (n-k)/h \rceil - [k = l \bmod h] \ge \frac{n}{h} - 2 \ge \frac{n}{2h}\,.$$

Now $J_{kl}(t) = j$ if and only if $Y_{kl}(t)$ has a certain value mod g. The case $n < 4h$ is trivial. $\square$

4. Uniformity. Let's assume now that, for given l and t, the random variables $J_{kl}(t)$ have a perfectly uniform distribution. Since the variables $J_{kl}(t)$ are independent for $0 \le k < h$, this means that

$$\Pr\left[J_{0l}(t) = j_0, J_{1l}(t) = j_1, \ldots, J_{(h-1)l}(t) = j_{h-1}\right] = \frac{1}{g^h} \tag{4.1}$$

for all h-tuples $(j_0, j_1, \ldots, j_{h-1})$.

In such a case the random variable $Q_l(t)$ defined in (2.3) is the sum of $h-1$ independent indicator variables, each equal to 1 with probability d/g because H has d elements. Hence $Q_l(t)$ has the binomial distribution with parameters $(h-1, d/g)$, and it is equal to r with probability

$$\binom{h-1}{r}\left(\frac{d}{g}\right)^r\left(1-\frac{d}{g}\right)^{h-1-r}. \tag{4.2}$$

Let $W^*_{jj'l}(t)$ be the value of $W_{jj'l}(t)$ under the assumption of uniformity $\bigl($see (2.4)$\bigr)$. Thus $W^*_{jj'l}(t)$ is independent of t, and we let $W^*_{jj'l} = W^*_{jj'l}(t)$ in accordance with (2.5). Then

$$W^*_{jj'l} = \frac{1}{g}\sum_{r=0}^{h-1}\binom{h-1}{r}\left(\frac{d}{g}\right)^r\left(1-\frac{d}{g}\right)^{h-1-r}\left|r - \left\lfloor\frac{hd}{g}\right\rfloor\right|. \tag{4.3}$$

For given values of d and j, the index $j' = (j + hd) \bmod g$ is at distance d from j. Suppose $a(d)$ of these pairs (j, j') have $j < j'$. Then $g - a(d)$ of them have $j > j'$, and $a(g-d) = g - a(d)$ since j is at distance $g - d$ from j'. The sum of (4.3) over all $j < j'$ is therefore independent of $a(d)$:

$$\begin{aligned}
\sum_{0 \le j < j' < g} W^*_{jj'l} &= \sum_{d=1}^{g-1}\frac{a(d)}{g}\sum_{r=0}^{h-1}\binom{h-1}{r}\left(\frac{d}{g}\right)^r\left(1-\frac{d}{g}\right)^{h-1-r}\left|r - \left\lfloor\frac{hd}{g}\right\rfloor\right| \\
&= \sum_{d=1}^{g-1}\frac{a(g-d)}{g}\sum_{r=0}^{h-1}\binom{h-1}{h-1-r}\left(\frac{g-d}{g}\right)^{h-1-r}\left(1-\frac{g-d}{g}\right)^r \\
&\qquad\qquad \times\left|h-1-r-\left\lfloor\frac{h(g-d)}{g}\right\rfloor\right| \\
&= \frac{1}{2}\sum_{d=1}^{g-1}\sum_{r=0}^{h-1}\binom{h-1}{r}\left(\frac{d}{g}\right)^r\left(1-\frac{d}{g}\right)^{h-1-r}\left|r - \left\lfloor\frac{hd}{g}\right\rfloor\right|.
\end{aligned}$$

(We have used the fact that $\lfloor h(g-d)/g\rfloor = h - 1 - \lfloor hd/g\rfloor$ when hd/g is not an integer.) But this is just the quantity $\psi(h,g)$ in (0.2), for each value of l. We have proved

Lemma 4. *If we assume that the variables $J_{kl}(t)$ have exactly the uniform distribution, the quantity (2.6) is exactly $\psi(h,g)n$.*

5. Perturbation. To complete the proof of (0.3), we use a general technique applicable to the analysis of many algorithms: If a given complicated algorithm $\mathcal{A}$ almost always has the same performance characteristics as a simpler algorithm $\mathcal{A}^*$, then the expected performance of $\mathcal{A}$ is the same as the performance of $\mathcal{A}^*$ plus an error term based on the cases where $\mathcal{A}$ and $\mathcal{A}^*$ differ. (See, for example, the analysis in [8], where this "principle of negligible perturbation" is applied to a nontrivial branching process.)

In the present situation we retain the $(n-1)$-dimensional probability space $(X_0, \ldots, X_{l-1}, t, X_{l+1}, \ldots, X_{n-1})$ on which the random variables $J_{kl}(t)$ were defined in (2.2), and we define a new set of random variables $J^*_{kl}(t)$ on the same space, where $J^*_{kl}(t)$ has exactly a uniform distribution on $\{0, 1, \ldots, g-1\}$. This can be done in such a way that $J_{kl}(t) = J^*_{kl}(t)$ with high probability.

More precisely, when l and t are given, $J_{kl}(t)$ depends only on the variables $X_{k'}$ with $k' \equiv k \pmod{h}$ and $k' \neq l$. The unit cube on these variables is partitioned into g parts $P_0, P_1, \ldots, P_{g-1}$ such that $J_{kl}(t) = j$ when the variables lie in P_j; the volume of P_j is $\Pr[J_{kl}(t) = j]$. We will divide each P_j into g sets P'_{j0}, P'_{j1}, $\ldots$, $P'_{j(g-1)}$, and define $J^*_{kl}(t) = i$ on P'_{ji}. This subdivision, performed separately for each k, will yield independent random variables $J^*_{0l}(t)$, $J^*_{1l}(t)$, $\ldots$, $J^*_{(h-1)l}(t)$. We will show that the subdivision can be done in such a way that

$$\Pr[J^*_{kl}(t) = j] = 1/g\,, \tag{5.1}$$

$$\Pr[J^*_{kl}(t) \neq J_{kl}(t)] < \phi(t)\,, \tag{5.2}$$

for $0 \leq j < g$ and $0 \leq k < h$. Thus, we will have perturbed the values of $J_{kl}(t)$ with low probability when $\phi(t)$ is small.

The following construction does what we need, and more:

Lemma 5. *Let $p_1, \ldots, p_m$ and $p^*_1, \ldots, p^*_m$ be nonnegative real numbers with $p_1 + \cdots + p_m = p^*_1 + \cdots + p^*_m = 1$. Then there are nonnegative reals p'_{ij} for $1 \leq i, j \leq m$ such that*

$$p_i = \sum_{j=1}^{m} p'_{ij}\,, \tag{5.3}$$

$$p^*_j = \sum_{i=1}^{m} p'_{ij}\,, \tag{5.4}$$

and

$$\sum_{i \neq j} p'_{ij} = 1 - \sum_j p'_{jj} = \frac{1}{2} \sum_j |p_j - p^*_j|. \tag{5.5}$$

Proof. This is a special case of "maximal coupling" in probability theory [5; 9, §III.14]; it can be proved as follows.

Let $p'_{jj} = \min(p_j, p^*_j)$, and observe that

$$\sum_j p'_{jj} = \sum_j \min(p_j, p^*_j) = \sum_j \tfrac{1}{2}(p_j + p^*_j - |p_j - p^*_j|) = 1 - \frac{1}{2} \sum_j |p_j - p^*_j|\,. \tag{5.6}$$

The existence of nonnegative p'_{ij}, $i \neq j$, such that (5.3) and (5.4) hold follows from the max flow–min cut theorem [4]: Consider a network with a source s, a sink t, and $2m$ nodes $v_1, \ldots, v_m, v^*_1, \ldots, v^*_m$; the edges are sv_j with capacity $p_j - p'_{jj}$, $v^*_j t$ with capacity $p^*_j - p'_{jj}$, and $v_i v^*_j$ with infinite capacity. $\square$

6. The effect of perturbation. When independent random variables $J^*_{kl}(t)$ have been defined satisfying (5.1) and (5.2), we can use them to define $Q^*_l(t)$ as in (2.3) and $W^*_{jj'l}(t)$ as in (2.4). This value $W^*_{jj'l}(t)$ has already been evaluated in (4.3); we want now to use the idea of perturbation to see how much $W_{jj'l}(t)$ can differ from $W^*_{jj'l}(t)$.

Since $Q_l(t) = O(h)$ and

$$|Q_l(t) - Q^*_l(t)| \le \sum_{k=0}^{h-1} [J_{kl}(t) \ne J^*_{kl}(t)]\,, \tag{6.1}$$

we have

$$\begin{aligned}\bigl|W_{jj'l}(t) - W^*_{jj'l}(t)\bigr| &= \Bigl|\bigl(\Pr[J_{ll}(t)=j] - \Pr[J^*_{ll}(t)=j]\bigr)\,\mathrm{E}\,\bigl|Q_l(t) - \lfloor hd/g\rfloor\bigr| \\ &\qquad + \Pr[J^*_{ll}(t)=j]\bigl(\mathrm{E}\,\bigl|Q_l(t)-\lfloor hd/g\rfloor\bigr| - \mathrm{E}\,\bigl|Q^*_{kl}(t) - \lfloor hd/g\rfloor\bigr|\bigr)\Bigr| \\ &< \frac{1}{g}\,\phi(t)O(h) + \frac{1}{g}\sum_{k=0}^{h-1}\Pr[J_{kl}(t)\ne J^*_{kl}(t)] \\ &= O\left(\frac{h}{g}\right)\phi(t)\,. \end{aligned} \tag{6.2}$$

(We assume that $J^*_{kl}(t) = J^*_{(k \bmod h)l}(t)$ when $k \ge h$.)

To complete our estimate we need to integrate this difference over all t.

Lemma 6. $\int_0^1 \phi(t)\,dt = O(g^2h/n)$.

Proof. The case $n < 4h$ is trivial. Otherwise we have

$$\begin{aligned}\int_0^1 \phi(t)\,dt &= 2\int_0^{1/2}\phi(t)\,dt \\ &< 4\int_0^{1/2}\sum_{k=1}^{\infty} e^{-2tk^2n/g^2h}\,dt \\ &< 4\int_0^{\infty}\sum_{k=1}^{\infty} e^{-2tk^2n/g^2h}\,dt \\ &= 4\sum_{k=1}^{\infty}\frac{g^2h}{2k^2n} = \frac{\pi^2}{3}\,\frac{g^2h}{n}\,. \quad \square\end{aligned}$$

Theorem 1. *The average number of inversions remaining after h-sorting and then g-sorting a random permutation of n elements, when h is relatively prime to g, is $\psi(h,g)n + O(g^3h^2)$, where $\psi(h,g)$ is given by (0.2).*

Proof. By (6.2) and Lemmas 2, 4, and 6, the average is $\psi(h,g)n$ plus

$$\begin{aligned}\sum_{\substack{0\le j<j'<g\\0\le l<n}}\int_0^1\bigl(W_{jj'l}(t) - W^*_{jj'l}(t)\bigr)\,dt &= O(g^2n)O(h/g)\int_0^1\phi(t)\,dt \\ &= O(g^3h^2)\,. \quad \square\end{aligned}$$

Notice that the proof of this theorem implicitly uses Lemma 5 for each choice of l and t, without requiring any sort of continuity between the values of $J^*_{kl}(t)$ as t varies. We could have defined $J^*_{kl}(t)$ in a continuous fashion; indeed, the random variables $[X_k < t]$ partition the $(n-1)$-cube into 2^{n-1} subrectangles in each of which $J_{kl}(t)$ has a constant value, so we could define $J^*_{kl}(t)$ over $(n-1)$-dimensional rectangular prisms with smooth transitions as a function of t. But such complicated refinements are not necessary for the validity of the perturbation argument.

7. Asymptotics. Our next goal is to estimate $\psi(h,g)$ when h and g are large. Notice that

$$\psi(h,g) = \frac{1}{2}\sum_{d=1}^{g-1} \mathrm{E}\left|Z(h-1,d/g) - \left\lfloor\frac{hd}{g}\right\rfloor\right| \tag{7.1}$$

where $Z(m,p)$ has the binomial distribution with parameters m and p. The mean of $Z(h-1,d/g)$ is $(h-1)d/g = \lfloor hd/g\rfloor + O(1)$, and the variance is $(h-1)d(g-d)/g^2$. If we replace Z by a normally distributed random variable with this same mean and variance, the expected value of $|Z - \lfloor hd/g\rfloor|$ is approximately $(2\pi)^{-1/2}\int_{-\infty}^{\infty}|t|e^{-t^2/2}\,dt = 2/\sqrt{2\pi}$ times the standard deviation, so (7.1) will be approximately

$$\frac{1}{g}\sqrt{\frac{h}{2\pi}}\sum_{d=1}^{g-1}\sqrt{d(g-d)}\,. \tag{7.2}$$

The detailed calculations in the remainder of this section justify this approximation and provide a rigorous error bound.

Lemma 7. *If Z has the binomial distribution with parameters (m,p), and $\lfloor mp\rfloor \le a \le \lceil mp\rceil$, then*

$$\mathrm{E}\,|Z-a| = \sqrt{\frac{2p(1-p)m}{\pi}} + O\Big(\frac{1}{\sqrt{mp(1-p)}}\Big)\,. \tag{7.3}$$

Proof. Consider first the case $a = mp$. By a formula of De Moivre [1, page 101] and Poincaré [10, pages 56–60], see Diaconis and Zabell [2],

$$\mathrm{E}\,|Z-mp| = 2\lceil mp\rceil\binom{m}{\lceil mp\rceil}p^{\lceil mp\rceil}(1-p)^{m+1-\lceil mp\rceil}. \tag{7.4}$$

In order to prove (7.3) in this case we may assume that $p \le 1/2$, since $|Z-mp| = |m-Z-m(1-p)|$. Moreover, we may assume that $mp > 1$ since (7.3) otherwise is trivial. Then, a routine application of Stirling's approximation shows that

$$\mathrm{E}\,|Z-mp| = \sqrt{\frac{2p(1-p)m}{\pi}}\exp\left(O\Big(\frac{1}{mp}\Big)\right)\,. \tag{7.5}$$

Next observe that if $\lfloor mp\rfloor \le a \le \lceil mp\rceil$, we have

$$\mathrm{E}\,|Z-a| = \mathrm{E}\,|Z-mp| + (mp-a)\big(1 - 2\Pr[Z \le mp]\big)\,. \tag{7.6}$$

Since $\Pr[Z\le mp] = \frac{1}{2} + O\big((mp(1-p))^{-1/2}\big)$, for example by the Berry–Esseen estimate of the error in the central limit theorem [3, §XVI.5], the result follows. $\square$

Corollary. *The asymptotic value of* $\psi(h,g)$ *is*

$$\psi(h,g) = \sqrt{\frac{\pi h}{128}}\, g + O(g^{-1/2}h^{1/2}) + O(gh^{-1/2})\,. \tag{7.7}$$

Proof. Since $\lfloor hd/g \rfloor \le \lfloor (h+1)d/g \rfloor \le \lfloor (hd+g-1)/g \rfloor = \lceil hd/g \rceil$, Lemma 7 yields

$$\begin{aligned}\psi(h+1,g) &= \frac{1}{2}\sum_{d=1}^{g-1} \mathrm{E}\,\left| Z(h,d/g) - \left\lfloor \frac{(h+1)d}{g} \right\rfloor \right| \\ &= \sum_{d=1}^{g-1}\left(\sqrt{\frac{h}{2\pi}\frac{d}{g}\Bigl(1-\frac{d}{g}\Bigr)} + O\left(\left(h\frac{d}{g}\Bigl(1-\frac{d}{g}\Bigr) \right)^{-1/2} \right) \right) \\ &= \sqrt{\frac{h}{2\pi}} \sum_{d=1}^{g-1} \sqrt{\frac{d}{g}\Bigl(1-\frac{d}{g}\Bigr)} + O(gh^{-1/2})\,.\end{aligned}$$

And Euler's summation formula with $f(x) = \sqrt{(x/g)(1-x/g)}$ tells us that

$$\begin{aligned}\sum_{d=1}^{g-1} f(d) &= \int_1^{g-1} f(x)\,dx + \frac{1}{2}f(1) + \frac{1}{2}f(g-1) + \frac{1}{12}f'(g-1) - \frac{1}{12}f'(1) - R \\ &= g\int_0^1 \sqrt{t(1-t)}\,dt + O(g^{-1/2}) = \frac{\pi g}{8} + O(g^{-1/2})\end{aligned}$$

because

$$|R| = \left| \int_1^{g-1} \frac{B_2(x \bmod 1)}{2} f''(x)\,dx \right| \le \frac{1}{12}\int_1^{g-1} \bigl|f''(x)\bigr|\,dx = \frac{1}{12}f'(1) - \frac{1}{12}f'(g-1)\,. \qquad \square$$

The error term is thus $O(g^{1/2})$ when $h = g^2+1$; for example, we have

h	g	$\psi(h,g)$	$\sqrt{\pi h/128}\,g$	difference/$\sqrt{g}$
901	30	140.018	141.076	0.1933
1601	40	249.539	250.741	0.1900
2501	50	390.412	391.739	0.1877

8. Common factors. Now let's consider the behavior of shellsort with increments $(ch, cg, 1)$, where c is an integer > 1. It is easy to see that the first two passes are equivalent to the first two passes of $(h,g,1)$ shellsort on c independent subarrays $(X_a, X_{a+c}, X_{a+2c}, \ldots)$ of size $\lceil (n-a)/c \rceil$ for $0 \le a < c$. The inversions that remain are the $\psi(h,g)n + O(g^3h^2c)$ inversions within these subarrays, plus "cross-inversions" between $\binom{c}{2}$ pairs of subarrays.

Yao [11, Theorem 2] proved that the average number of cross-inversions is $\frac{1}{8}\sqrt{\pi c}\,(1-c^{-1})n^{3/2} + O(cghn)$. The following lemma improves his error term slightly.

Lemma 8. *The average number of cross-inversions after ch-sorting and cg-sorting is*

$$\frac{1}{8}\sqrt{\pi c}\,(1-c^{-1})n^{3/2}+O(cgh^{1/2}n)+O(c^2g^3h^2)\,. \tag{8.1}$$

Proof. Let's consider first the process of h-sorting and g-sorting two independent arrays $(X_0, X_1, \ldots, X_{n-1})$ and $(\widehat{X}_0, \widehat{X}_1, \ldots, \widehat{X}_{n-1})$, then interleaving the results to obtain $(X_0'', \widehat{X}_0'', X_1'', \widehat{X}_1'', \ldots X_{n-1}'', \widehat{X}_{n-1}'')$. The cross inversions are then the pairs $\{X_l'', \widehat{X}_{l'}''\}$ where either $X_l'' > \widehat{X}_{l'}''$ and $l \le l'$ or $X_l'' < \widehat{X}_{l'}''$ and $l > l'$.

Recasting this process in the model of section 2 above, we assume that $X_l = t$, while the other $2n-1$ variables $(X_0, \ldots, X_{l-1}, \ldots, X_{n-1}, \widehat{X}_0, \ldots, \widehat{X}_{n-1})$ are independent and uniformly distributed between 0 and 1. We define

$$Y_{kl}(t) = \sum_{\substack{k' \equiv k \pmod{h} \\ 0 \le k' < n}} [X_{k'} < t]\,, \quad \widehat{Y}_{kl}(t) = \sum_{\substack{k' \equiv k \pmod{h} \\ 0 \le k' < n}} [\widehat{X}_{k'} < t] \tag{8.2}$$

as in (2.1). The elements of each array are divided into h subarrays by h-sorting, and the elements $< t$ have $Y_{kl}(t)$ and $\widehat{Y}_{kl}(t)$ elements in the kth subarrays. Then g-sorting will form g lists, with

$$L_{jl}(t) = \sum_{k=0}^{h-1} \left\lceil \frac{Y_{kl}(t) - a_{kj}}{g} \right\rceil \tag{8.3}$$

elements $< t$ in the jth list of the first array, where $a_{kj} \in \{0, 1, \ldots, g-1\}$ is given by $k + a_{kj}h \equiv j \pmod{g}$. Similarly, there will be

$$\widehat{L}_{jl}(t) = \sum_{k=0}^{h-1} \left\lceil \frac{\widehat{Y}_{kl}(t) - a_{kj}}{g} \right\rceil \tag{8.4}$$

elements $< t$ in the jth list of the second. Element $X_l = t$ of the first array will go into list $j = J_{ll}(t)$ as before, where $J_{kl}(t)$ is defined in (2.2). The number of cross-inversions between this element and the elements of the second array will then be

$$V_l(t) = \sum_{j'=0}^{g-1} \left| \widehat{L}_{j'l}(t) - L_{jl}(t) - [j' < j] \right|. \tag{8.5}$$

The average total number of cross-inversions is the sum of $\mathrm{E}\, V_l(t)$ over all l, integrated for $0 \le t \le 1$.

We know from Lemma 3 that the numbers $Y_{kl}(t) \bmod g$ have approximately a uniform distribution. Therefore

$$\left\lceil \frac{Y_{kl}(t) - a_{kj}}{g} \right\rceil = \frac{Y_{kl}(t) - a_{kj} + R_{jkl}(t)}{g}$$

where $R_{jkl}(t)$ is approximately uniform on $\{0, 1, \ldots, g-1\}$. It follows that

$$L_{jl}(t) = \frac{Z_l(t)}{g} + \sum_{k=0}^{h-1} \left(\frac{R_{jkl}(t) - a_{kj}}{g} \right), \tag{8.6}$$

where

$$Z_l(t) = \sum_{k=0}^{h-1} Y_{kl}(t)$$

is the total number of elements in the first array that are $< t$.

Since $R_{jkl}(t)$ depends on $Y_{kl}(t) \bmod g$ only, or equivalently on $J_{kl}(t)$, we may use the perturbed truly uniform random variables $J^*_{kl}(t)$ in section 5 (or repeat the argument there with $R_{jkl}(t)$) and construct random variables $R^*_{jkl}(t)$ that are uniform on $\{0, 1, \ldots, g-1\}$ and satisfy $\Pr[R^*_{jkl}(t) \neq R_{jkl}(t)] < \phi(t)$; moreover, the variables $R^*_{jkl}(t)$ are independent for $0 \le k < h$ and fixed j and l. Consequently

$$\mathrm{E}\,|R^*_{jkl}(t) - R_{jkl}(t)| \le g \Pr[R^*_{jkl}(t) \neq R_{jkl}(t)] < g\phi(t)\,. \tag{8.7}$$

By independence and the fact that $\mathrm{E}\,R^*_{jkl}(t) = (g-1)/2$,

$$\mathrm{E}\left(\sum_{k=0}^{h-1} R^*_{jkl}(t) - h(g-1)/2\right)^2 = \sum_{k=0}^{h-1} \mathrm{E}\left(R^*_{jkl}(t) - (g-1)/2\right)^2 < hg^2,$$

which by the Cauchy–Schwarz inequality yields

$$\mathrm{E}\left|\sum_{k=0}^{h-1} R^*_{jkl}(t) - h(g-1)/2\right| < \sqrt{h}g. \tag{8.8}$$

Let $W_{jl} = \frac{1}{g}\bigl(\sum_{k=0}^{h-1} R_{jkl}(t) - h(g-1)/2\bigr)$ and $b_j = \frac{1}{g}\bigl(h(g-1)/2 - \sum_{k=0}^{h-1} a_{kj}\bigr)$; then

$$L_{jl}(t) = \frac{Z_l(t)}{g} + W_{jl} + b_j, \tag{8.9}$$

where by (8.7) and (8.8)

$$\mathrm{E}\,|W_{jl}(t)| < \sqrt{h} + h\phi(t).$$

A similar argument shows that

$$\widehat{L}_{jl}(t) = \frac{\widehat{Z}_l(t)}{g} + \widehat{W}_{jl} + b_j.$$

Hence

$$V_l(t) = \sum_{j'=0}^{g-1}\left(\frac{|\widehat{Z}_l(t) - Z_l(t)|}{g} + O(|W_{j'l}| + |\widehat{W}_{j'l}| + 1)\right)$$

and

$$\mathrm{E}\,V_l(t) = \mathrm{E}\,|\widehat{Z}_l(t) - Z_l(t)| + O(g\sqrt{h}\,) + O(gh)\,\phi(t)\,. \tag{8.10}$$

The quantity $|\widehat{Z}_l(t) - Z_l(t)|$ is just what we would get if we were counting the cross-inversions between two fully sorted arrays that have been interleaved. Therefore

$$\int_0^1 \sum_{l=0}^{n-1} \mathrm{E}\left|\widehat{Z}_l(t) - Z_l(t)\right| dt$$

must be the average number of inversions of a random 2-ordered permutation of $2n$ elements; this, according to Douglas H. Hunt in 1967, is exactly $n2^{2n-2}/\binom{2n}{n}$ [7, exercise 5.2.1–14]. Since $\binom{2n}{n} = \bigl(1+O(1/n)\bigr)4^n/\sqrt{\pi n}$, we obtain the desired total

$$\int_0^1 \mathrm{E} \sum_{l=0}^{n-1} V_l(t)\, dt = \frac{\sqrt{\pi}\, n^{3/2}}{4} + O(gh^{1/2}n) + O(g^3h^2) \tag{8.11}$$

by Lemma 6. Similarly, the same result holds for two arrays of different sizes $n + O(1)$.

Lemma 8 follows if we replace n by $n/c + O(1)$ in (8.11) and multiply by $\binom{c}{2}$. □

9. The total cost. So far we have been considering only the number of inversions removed during the third pass of a three-pass shellsort. But the first two passes can be analyzed as in Yao's paper [11]:

Theorem 2. *Let g and h be relatively prime and let c be a positive integer. The average number of inversions removed when $(ch, cg, 1)$-shellsort is applied to a random n-element array is*

$$\frac{n^2}{4ch} + O(n) \tag{9.1}$$

on the first pass,

$$\frac{1}{8g}\sqrt{\frac{\pi}{ch}}\,(h-1)n^{3/2} + O(hn) \tag{9.2}$$

on the second, and

$$\psi(h,g)n + \frac{1}{8}\sqrt{\frac{\pi}{c}}\,(c-1)n^{3/2} + O\bigl((c-1)gh^{1/2}n\bigr) + O(c^2g^3h^2) \tag{9.3}$$

on the third.

Proof. The first pass removes an average of $\frac{1}{4}\bigl(n/ch + O(1)\bigr)^2$ inversions from ch subarrays of size $\lfloor n/ch \rfloor$ or $\lceil n/ch \rceil$; this proves (9.1). The second pass is equivalent to the second pass of $(h, g, 1)$-shellsort on c independent subarrays of sizes $\lfloor n/c \rfloor$ or $\lceil n/c \rceil$. Equation (9.3) is Lemma 8. So the theorem will follow if we can prove (9.2) in the case $c = 1$. And that case follows from [11, equation (32)], with the $O(n)$ term replaced by $O(n/kh)$ in the notation of that paper. (See also [7, second edition, exercise 5.2.1–40.) □

Corollary. *If $h = \Theta(n^{7/15})$, $g = \Theta(n^{1/5})$, and $\gcd(g,h) = 1$, the running time of $(h, g, 1)$-shellsort is $O(n^{23/15})$.*

Proof. The first pass takes time $O(n^{2-7/15})$, by (9.1); the second takes $O(n^{3/2+7/30-1/5}) + O(n^{1+7/15})$, by (9.2); and the third takes $O(n^{1+1/5+7/30}) + O(n^{3/5+14/15})$ by (7.6) and (9.3). □

10. Two conjectures. Our estimate $O(g^3h^2)$ for the difference between $\psi(h,g)n$ and the average number of third-pass inversions may not be the best possible. In fact, the authors conjecture that the difference is at most $O(g^3h^{3/2})$. This sharper bound may perhaps follow from methods analogous to those in the proof of Lemma 8.

If such a conjecture is valid, the running time of $(h,g,1)$-shellsort will be $O(n^{3/2})$ when $h \approx n^{1/2}$ and $g \approx n^{1/4}$. A computer program was written to test this hypothesis by applying $(h,g,1)$-shellsort to random arrays of n elements with $h = g^2+1$ and $n = g^2h = g^4+g^2$. The following empirical results were obtained, to three significant figures:

g	inversions	$\psi(h,g)n$	g	inversions$/10^5$	$\psi(h,g)n/10^5$
1	0 ± 0	0	17	$36.6 \pm 2.36/32$	37.3
2	$7.12 \pm 2.09/100$	7.5	18	$51.7 \pm 3.35/32$	52.6
3	$94.4 \pm 13.6/100$	98.3	19	$71.5 \pm 4.81/32$	72.9
4	$563 \pm 59.1/100$	581	20	$97.3 \pm 6.14/10$	99.2
5	$2210 \pm 195/100$	2280	21	$130 \pm 8.93/10$	133
6	$6740 \pm 560/100$	6910	22	$174 \pm 12.3/10$	176
7	$17200 \pm 1300/100$	17600	23	$226 \pm 14.0/10$	230
8	$38600 \pm 2820/100$	39500	24	$291 \pm 16.8/10$	297
9	$78900 \pm 5670/100$	80600	25	$368 \pm 23.7/10$	380
10	$149000 \pm 10600/100$	152000	26	$475 \pm 29.1/10$	480
11	$265000 \pm 17200/32$	271000	27	$595 \pm 39.0/10$	603
12	$447000 \pm 30300/32$	458000	28	$735 \pm 44.9/10$	750
13	$727000 \pm 49300/32$	742000	29	$922 \pm 52.1/10$	926
14	$1140000 \pm 75400/32$	1160000	30	$1110 \pm 74.0/10$	1140
15	$1730000 \pm 116000/32$	1760000	31	$1370 \pm 97.9/10$	1380
16	$2530000 \pm 166000/32$	2590000	32	$1650 \pm 101/10$	1670

(The inversion counts are given here in the form $\mu \pm \sigma/\sqrt{r}$, where μ and σ are the empirical mean and standard derivation in r independent trials. For example, 10000 trials were made when $g \le 10$, but only 100 trials were made when $g \ge 20$.) Both mean and standard derivation seem to be growing proportionately to $g^6 \approx n^{3/2}$, with $\sigma \approx \mu/15$ for $g \ge 10$.

These data suggest also another conjecture, that the average number of inversions is $\le \psi(h,g)n$ when h and g are relatively prime. Indeed, the deviations from uniformity between $\mathcal{A}$ and $\mathcal{A}^*$ should tend to cause fewer inversions, because $\mathcal{A}$ forces the balance condition $Y_{kl}(1) = n/h + O(1)$ for all k and l. This second conjecture obviously implies running time $\Theta(n^{3/2})$ when $h = \Theta(n^{1/2})$ and $g = \Theta(n^{1/4})$.

11. More than three increments? It may be possible to extend this analysis to $(h,g,f,1)$-shellsort, by analyzing the following stochastic algorithm. "Initialize two sets of counters $(I_0, I_1, \ldots, I_{g-1})$ and $(J_0, J_1, \ldots, J_{h-1})$ by setting $I_j \leftarrow j \bmod f$ and $J_k = k \bmod g$ for all j and k. Then execute the following procedure n times: Choose a random k in the range $0 \le k < h$. Set $j \leftarrow J_k$ and $i \leftarrow I_j$; then set $J_k \leftarrow (J_k + h) \bmod g$ and $I_j \leftarrow (I_j + g) \bmod f$."

Consider the transition from $l = l_t$ to $l' = l_{t+1}$ in the proof of Lemma 1. When elements enter the array in increasing order, the choice of k represents the subarray that will contain a new element X during the h-sort; then X goes into list j during the g-sort, and into list i during the f-sort. We can therefore obtain the contribution of X to the inversions between lists i and i' for $i < i' < f$, by considering a state P_l obtained from the I table just as Q_l was obtained from the J table in Lemma 1.

References

[1] Abraham De Moivre, *Miscellanea Analytica de Seriebus et Quadraturis*, (London: J. Tonson and J. Watts, 1730).

[2] Persi Diaconis and Sandy Zabell, "Closed form summation for classical distributions: Variations on a theme of De Moivre," *Statistical Science* **6** (1991), 284–302.

[3] William Feller, *An Introduction to Probability Theory and Its Applications* **2** (New York: Wiley, 1966).

[4] L. R. Ford, Jr., and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics* **8** (1956), 399–404.

[5] Sheldon Goldstein, "Maximal coupling," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **46** (1979), 193–204.

[6] Donald E. Knuth, *Seminumerical Algorithms*, Volume 2 of *The Art of Computer Programming* (Reading, Massachusetts: Addison–Wesley, 1969). Second edition, 1981.

[7] Donald E. Knuth, *Sorting and Searching*, Volume 3 of *The Art of Computer Programming* (Reading, Massachusetts: Addison–Wesley, 1973). Second edition, 1997.

[8] Donald E. Knuth, Rajeev Motwani, and Boris Pittel, "Stable husbands," *Random Structures and Algorithms* **1** (1990), 1–14.

[9] Torgny Lindvall, *Lectures on the Coupling Method* (New York: Wiley, 1992).

[10] Henri Poincaré, *Calcul des Probabilités* (Paris: Georges Carré, 1896).

[11] Andrew Chi-Chih Yao, "An analysis of $(h, k, 1)$-Shellsort," *Journal of Algorithms* **1** (1980), 14–50.

Authors' addresses:

Svante Janson, Department of Mathematics, Uppsala University, P.O.Box 480, 75106 Uppsala, Sweden; `svante.janson@math.uu.se`

Donald E. Knuth, Computer Science Department, Gates Building 4B, Stanford University, Stanford CA 94305–9045 USA; `http://www-cs-faculty.stanford.edu/~knuth`